

WWW Image Searching Delivers High Precision and No Misinformation: Reality or Ideal?

Paul Nieuwenhuysen
Vrije Universiteit Brussel, Brussel, Belgium

Paul.Nieuwenhuysen@vub.ac.be

Abstract

On the one hand information storage and retrieval through the Internet has made spectacular progress. On the other hand, searching for information still confronts us with information systems that are far from perfect. The purpose of this paper is to clarify the practical state of the art of finding information through image WWW search systems in a particular subject domain.

Queries in a particular domain of interest were submitted to public access WWW image search engines. The results were evaluated quantitatively. For each search, the scoring method leads to a positive score related to precision and to a negative score related to misinformation.

Most image search results lead to information that is not relevant, due to known imperfections in information retrieval. Furthermore, a significant number of results are even misleading. This led to the conclusion that information retrieval from the WWW through image searching is attractive, simple and fast. However, users have to live with less than ideal retrieval systems and should be careful with misinformation. Based on the outcomes of this investigation, recommendations are formulated to, authors, publishers, librarians, and end-users.

Keywords: information retrieval, image searching, Google, precision, misinformation, information literacy

Introduction

Information Access through the Internet

Advances in information science and technology have made available:

- a huge amount of information on the Internet, as well as
- search engines to discover relevant information that are user friendly and perform faster and better than ever.

This makes WWW searching the first method to find information for a growing number of users.

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact 0HPublisher@InformingScience.org to request redistribution permission.

A review of the state of the art is given by Manning, Raghavan, & Schutze, 2008.

On the other hand, many users are still frustrated by:

- scarcity of information that is available through the WWW, in the case of many specific information needs and

- low precision of WWW search results.

This author experiences this also more formally in years of guiding and teaching students in universities and scientific workshops.

Searching for Images on the WWW

WWW image search systems are of course mainly suitable to find relevant images on the WWW. It may be somewhat confusing that these systems do not search for images in some magical or real way, but that they are in fact text search engines that do not show texts with links as results but text search engines that yield thumbnail images with links to the original document on the WWW.

Moreover, image searching can also be used to find relevant documents by first selecting relevant images and by subsequent linking to the complete document. An advantage in comparison with more common text search engines is that in many cases images can be screened and selected faster than texts; we can summarize this as ‘seeing is faster than reading’.

For the past few years, the big companies that offer the most popular WWW search engines are Google, Microsoft (MSN, Live, Bing...) and Yahoo! All these offer also more dedicated image search engines (JISC, 2008).

Misinformation besides Information on the WWW

Content on the Internet is a mixture of:

- sound, valuable, neutral, scientific information that is offered in many cases by educational and non-profit organizations,
- information that is biased, often by commercial interests,
- useless documents without informative value,
- even “disinformation” or “misinformation”.

The presence of lower quality information makes it difficult and challenging for retrieval systems to provide high quality search results and for the user to find relevant information. This is important and has been investigated mainly in medicine (see the review by Eysenbach, Powell, Kuss, & Sa, 2002, and the more recent investigation by Ostry, Young, & Hughes, 2007).

The author of this paper focuses on a different subject domain, namely art and more specifically classical African sculptural art. In recent years while screening all kinds of documents related to African art, such as WWW pages, logos, brochures and even articles and books, I have observed that many show objects as “African art”, whereas people familiar with this field see on the photos objects of an embarrassingly low quality, objects that are based on or copied from real, old, classical, high quality, authentic African art. The authors are often students, but also others who may be well placed to create a logo or write a book, but who have clearly not much expertise in the area of African art. More specifically, I have observed in more than one such case the photos had simply been copied from the WWW, probably because this is convenient, fast and easy.

All this indicates that:

- finding appropriate images is still not straightforward
- besides information, also misinformation is distributed through the WWW
- information users can be misinformed by the content of WWW pages
- misinformed users become new sources of more misinformation.

4. a manual reproduction of
5. an original, primary object.

Assessing the impact of the recent media (r)evolution requires attention and research.

Problem Statements / Purpose of this Research

The facets of our reality today that are sketched in the Introduction above bring us to several problems that deserve investigation. A broad, general question is: What is the state of the art of information retrieval through WWW image searching in practice for common users, as seen from the user's point of view? More concrete research problems are:

- How do the various leading, popular WWW image search engines perform in aspects such as user interface, coverage of information sources, recall, precision and relevance ranking of search results? In particular how do these systems perform for a common user without exceptional, advanced skills in information retrieval?
- In which way does the amalgamation of content and older media in the newer digital media influence discovery and appreciation of content such as works of art? More specifically, in which way and to which extent plays misinformation a role in the process and in the experience of the user, besides the more useful content and information?

The outcomes of the investigation should lead to recommendations for authors, publishers, librarians and end-users of information systems.

In practice, we have executed searches typical for common users and we have analyzed the results.

Methods

Preliminary investigations have served to refine the procedures and have lead to the following methods.

WWW Searching for Images

As Internet browser Mozilla Firefox 3.5 and higher versions were used.

In Google Image search, "Safe search" and exploiting web history were set OFF and no restrictions in image size or type were made. Analogous settings were used with other image search engines.

In all search engines, all queries were formulated:

- in the way that most users apply a search engine, simply with few words, without synonyms, without quotation marks that could be used to indicate a search for a phrase composed of several words, and without explicit Boolean operators such as AND and OR,
- in agreement with the syntax applicable in the search engine
- using specific names of ethnic styles

Many common users formulate less efficient queries:

- Some users use synonyms in their query without Boolean OR operator; in that case the space between words is in most cases interpreted by the search engine as a Boolean AND, so that precision may increase, but recall may decrease dramatically so that only few or no relevant results show up.

- Users with limited knowledge about specific names and terms in African art will obtain result sets that suffer from a lower precision.

So the procedure applied in this investigation yields search results that give probably an optimistic, bright view on reality. In other words, the results do not reflect a worst case scenario.

For each query, the first 20 retrieved thumbnail images were inspected and taken into account for scoring. This is justified as follows:

- Google Image search shows thumbnails in groups of 20.
- Search engines attempt to rank results according to their relevance for the user, so that for most queries the relevance of further results is lower. In many cases with an interesting result set, further thumbnail images were inspected anyhow and this confirmed in every case that the precision of the results further down beyond the first 20 was indeed getting worse.
- This procedure corresponds with the way in which most common users apply the search engines.
- This approach makes it feasible to execute this investigation manually in practice.

The topic of all searches was chosen so that image searching is more efficient than normal web text searching to select suitable / relevant / interesting / promising results. More concretely, the topic of each search was a type of classical African art object (as explained below).

Each concrete search query that was submitted and the scores given to the results (as described below) were noted in one row of a spreadsheet on computer to analyze the data and to produce charts. A file with these data collected can be obtained from the author on request.

More than 150 queries on more than 35 types of objects have been executed up to now, with subsequent evaluation of the results. Examples are given in the Appendix.

Topics of the Searches: African Art Objects

The aim of each search was to find information on a type of object in the area of primitive, tribal, primal, traditional, classical, ethnic, old or antique, sculptural African art. Perhaps surprisingly, there is still no agreement on a brief, clear and neutral term to describe this subject domain. The adjectives “tribal” and “primitive” are criticized by many as bringing a negative connotation that should be avoided. In French, these days many use the term “arts premiers Africains” (for instance Martínez-Jacquet, 2009). Some books deal with worldwide so-called primitive art, including African art, e.g., Martínez-Jacquet (2009) and Miller (2006). Many books provide an introduction to old African art, including the following recent ones: Bacquart, 2002; Bassani, 2004, 2005; Carini, 2004; Eisenhofer & Guggei, 2002; Grootaers & Eisenburger, 2002; Herreman et al., 2003; Petridis, 2003; Pierrat, 2008; Robbins & Ingram Nooter, 2004; Stepan, 2001; Visona, Poynor, Cole, & Blier, 2007. Some books have a more limited scope and deal with African masks as artistic, cultural expressions, such as the following recent ones: Bouttiaux, 2009; Hahner, Vajda, & Kecskesi, 2007; Maertens, Neerman, & Burssens, 2009; Stepan & Hahner, 2005; Weinhold, 2001.

Recently, in October 2009, a fee-based database has become available about high quality African art objects that have been offered for sale by prestigious auction houses, <http://www.tribalindex.com/> Each database record offers information about one object: images, estimated value, selling price and so on, in as far as that information could be derived from available documents.

Interest in African art has been steadily growing as indicated these days by:

- an increasing number of books published on the topic,
- the production of the new specialized database on African art, and
- the annual, international fairs on tribal, ethnic art including old Oceanic and African art, in Brussels (named BRUNEAF <http://www.bruneaf.com/>), Paris (named Parcours des Mondes <http://www.parcours-des-mondes.com/>), Amsterdam (named Tribal Art Fair = TAF <http://www.tribal-art-fair.nl/>), London, San Francisco...

Evaluation and Scoring of Search Results

To make this investigation feasible in practice, we simplify reality and assign to each result set one positive and one negative score, as follows:

Positive score +1

The image shows antique, old, real, authentic, subtle, expressive objects of high quality, which have been made for use in the region where they were produced. Such objects are mainly shown by:

- a dealer offering at least a few high quality objects,
- a high end auction house such as Bonhams, Christie's, Koller, Lempertz, Neumeister, Pierre Bergé & Associés, Quittenbaum, Sotheby's,
- a museum,
- a university,
- an expert collector

Positive score +0.5

- The thumbnail image in the search results list is by definition small but it looks relevant and informative on its own, while linking to the original source does not yield more information. For instance the original source leaves us in doubt about the quality of the object shown or does not allow us to evaluate the quality or shows an object of medium quality. Sources of this kind of information are mainly medium quality dealers and collectors. Such results can be useful as they can finally yield a relevant text in spite of the low or medium or uncertain quality of the object. Even a photo of a medium quality object can reveal an artistic element that is worth to be discovered, as Picasso answered when he was told that the African art objects in his personal collection were not all of high quality.
- The thumbnail image looks relevant but leads only to the bibliographical description of a book that may be interesting but that is not directly available through the WWW.
- The retrieved thumbnail image does not show a relevant picture, but nevertheless linking to the full document yields relevant information. (Of course only few links are further investigated when they do not show a relevant image and in practice a common user does not want to spend time in this way.)

On the other hand, such results can also be considered partly as a waste of time for the user or even as misinformation, and therefore, besides the positive +0.5, also a negative score -0.5 is given.

Positive or negative neutral score 0

- The thumbnail image is clearly irrelevant, even when it shows a photo of an object that may be seen as relevant by a user with little knowledge about the subject.
- The thumbnail image looks promising / relevant / interesting, but the link to the image source does not lead to any information due to a broken link.
- The thumbnail image looks promising / relevant / interesting, but the link to the image source leads to a web page without useful relation to the image.

Negative score -0.5

- A negative score -0.5 is assigned in several cases in which also a positive score +0.5 is given, as described above.
- The thumbnail image may look relevant to a user, but shows a low quality item, and following the link leads to no further information or to a statement that makes clear to a user that the object is of low quality.
- The thumbnail leads to a site for which the browser gives a warning that the connection is “untrusted” or the system gives the message: “Malware warning: visiting this web site may harm your computer!”
- The thumbnail image looks promising, but leads to a page that shows only one or several low quality objects and the context makes it clear that the object is a reproduction, a copy, or that quality is low.

Negative score -1

- The source image shows an object of low quality, so-called fake, copy, airport art made for sale to tourists, to people from outside the tribe for decoration of their home and environment, that offers only a decadent or grotesque style, often with exaggerated typical features. Such objects are mainly shown by low-end dealers/sellers.

General remarks concerning the scores

The positive score is of course related to what is named in information science the “precision” of the information retrieval action. (Chapter 8 on Evaluation in information retrieval in Manning et al., 2008).

The negative score can be interpreted as an indication of misinformation.

In the case that retrieval gives several thumbnail images that link:

- to the same site, or
- to the same, duplicate or near-duplicate image or text information on sites residing on a different host computer,

then these thumbnails are taken into account only once. In other words only one positive and one negative score is assigned.

The author is well aware of the complexity of the concepts quality and authenticity of African sculptures, with many aspects, facets, and dimensions (see the well-known paper by Kamer, 1974). So the scoring method applied here should be seen as a practical simplification to start exploring this field of research.

The scores discussed above deliberately do not reflect aspects that could be taken into account in another type of quality score, but that we considered as less relevant and important in this investigation. Examples of such aspects are completeness, readability, design / clarity, and references provided (as mentioned by Eysenbach et al. 2002), plus accessibility, user-friendly architecture, speed of loading, image quality of the images shown, up-to-date information.

Value of African art objects

The scores that have been assigned to each search result have further been analyzed in relation to the market value of a typical high quality object that was the topic of the search query. To realize this in practice, values have been taken from data that have been collected over the years from the following sources:

- Books that give indications of price ranges of African art objects for dealers and collectors (Eisenhofer & Guggei, 2002; Miller, 2006), being aware of the fact that also such books can include not only authentic high quality items.
- Catalogs of classical live auctions, with value estimates, printed and on the WWW.
- Auction results, that is prices paid at a live auction, as found on the WWW.
- Auctions on the WWW, but these are mainly limited to low value items.
- Prices asked by dealers in shops, at sale exhibitions and on the WWW.

Recently a fee-based database has become available about high quality African art objects that have been offered for sale by prestigious auction houses, <http://www.tribalindex.com/>. This database can be useful for this kind of investigation. Nevertheless, this database was not used here for the following reasons:

- Access was not yet available when this investigation was already going on.
- The information is restricted to objects sold by only a few prestigious auction houses.
- The cost to access the database is not negligible.

Of course for each type of object a range of values exists. However, for a clear scatter chart, a range is not well applicable. Therefore sharp values have been assigned to each family of objects investigated, as follows:

- In one approach to create the scatter charts, the maximum value found was assigned, because this corresponds of course in most cases to a high quality object.
- However, such a high, maximum price can be artificial, exceptional and excessive. Therefore in another scatter chart, the second highest price found was displayed. Furthermore that value originates from a different dealer or auction house than the one that gave the maximum price.

A high accuracy for this variable is not meaningful, relevant or important in this context, because a strong, strict correlation between precision of search result sets and misinformation is not expected.

By using these two methods to bring value of objects into the analysis, we can check if different ways to assign values lead to views or conclusions that are significantly different or not.

Of course the value of objects changes over time, which again complicates the situation.

Results

Comparison of WWW Image Search Engines

In the first phase of this investigation, several WWW image search engines were evaluated and compared, by using them with the same queries. Concretely, the systems offered by the three leading companies in WWW searching were compared: these are already for several years Yahoo!, Microsoft, and Google (JISC, 2008). Since 2008, Microsoft has changed their search engines from Live to Bing. Several sources have announced in 2009 that Yahoo! is planning to cooperate more with Microsoft and to replace its own search system also by Bing.

For example, the query ‘Namji doll’ in Google Image Search leads mainly to low quality dolls inspired by high quality dolls created by the Namji / Dowayo people in Cameroon. This results in a very low positive score and in a very pronounced negative score. Repeating the query after a few days gave similar results. Executing the query in the new Bing Image Search gave a positive score of 0 and a negative score of 14; so this approach was even less satisfying.

In most cases the positive score that reflects precision was highest when Google was used. This agrees with the finding that Google offered the best WWW image search system in 2008 (JISC, 2008). Also Google offers the most popular search systems, as written in JISC (2008) and as can be concluded from the usage statistics collected for the WWW sites managed by myself.

Google Image Search shows in the results under each thumbnail image a well readable corresponding WWW site address that allows the user in most cases:

- to get an idea about the country or type of the web site (educational or commercial or...) and
- to see that some results come from the same web site.

In our experience this display method is more useful than the one offered by the other image search engines used.

In view of all this, the further investigation was carried out with Google WWW Image Search.

Up to now, more than 150 queries were submitted after which the results were analyzed. Each query aimed at one of more than 30 types of objects.

Recall through WWW Image Searching

To which extent can WWW image searching reveal relevant photos and documents on the WWW? Let us start with an example of a typical search experience. Information was searched about the typical wooden vertical pole sculptures of a man or woman that are created within the Tiv tribe in Nigeria and which are often named “ihambe”. Simple queries were used like ‘tiv ihambe’ or ‘ihambe’. Surprisingly, many thumbnail photos were retrieved of the famous singer and performer Michael Jackson who died recently, which are irrelevant for the search topic. Only 0 or 1 relevant image was found. These are only few results in comparison with more than 50 web pages/files/documents with photos that I have collected from the WWW, exploiting many sources with various methods to discover the information that is available in principle. This example is clear as the numbers involved are small and easy to handle in practice. Assessing recall quantitatively is less easy when numbers are higher. In any case, assessing the absolute value of recall in a retrieval system is hindered by:

- the need to know all the relevant items that are in principle available so that they should be found in an ideal system, and

- the need to investigate all search results and not only those that rank high in the result lists.

Anyway, the experience gained throughout the investigation that is reported further below, has shown that image searching is far from ideal to discover most relevant documents.

Automatic Query Expansion Decreases Precision

Google WWW text search as well as WWW image search includes in the executed query also spelling variations of the original query. This can of course increase recall, which may be desirable in some cases. However, in the context of this investigation it decreases the first 20 precision and therefore also the positive score assigned to the performance of the search system. For instance:

- The query for the Ethiopian “Kambatta” also yields irrelevant results that include the word “Khambatta” which is the name of a popular woman.
- The query for the people living in Burkina Faso named Mossi also yields irrelevant results that include the word Mossy.

Precision of Image Searches

A summary of the positive precision score of executed searches is presented as a bar chart in Figure 2. Each image query result set of 20 images is represented by a bar in this bar chart. More than 100 queries / bars are included. The length of each bar reflects the positive precision score.

The method applied to display the data makes it possible to read and interpret the chart as follows: A fully covered chart area would indicate an ideal information system, and a completely empty chart would correspond with a useless information system.

The actual chart mainly shows that precision scores are much lower than the maximum of 20. So the precision for this kind of searching is (still) far from ideal. The quality of the queries executed is better than expected in most real searches by common users, as outlined in the section on Methods. Nevertheless the median positive score was much lower than the maximum, namely 3 / 20.

This is partly due to known imperfections and stumbling blocks:

- in information retrieval in general, such as ambiguity of search terms, decreasing precision due to automatic expansion of the search query by the retrieval system (as illustrated above), and the lack of sound criteria to rank results,
- and more specifically in information retrieval from the WWW, such as results that lead to the same WWW site, broken links, and duplicate or very similar items retrieved from various WWW sites.

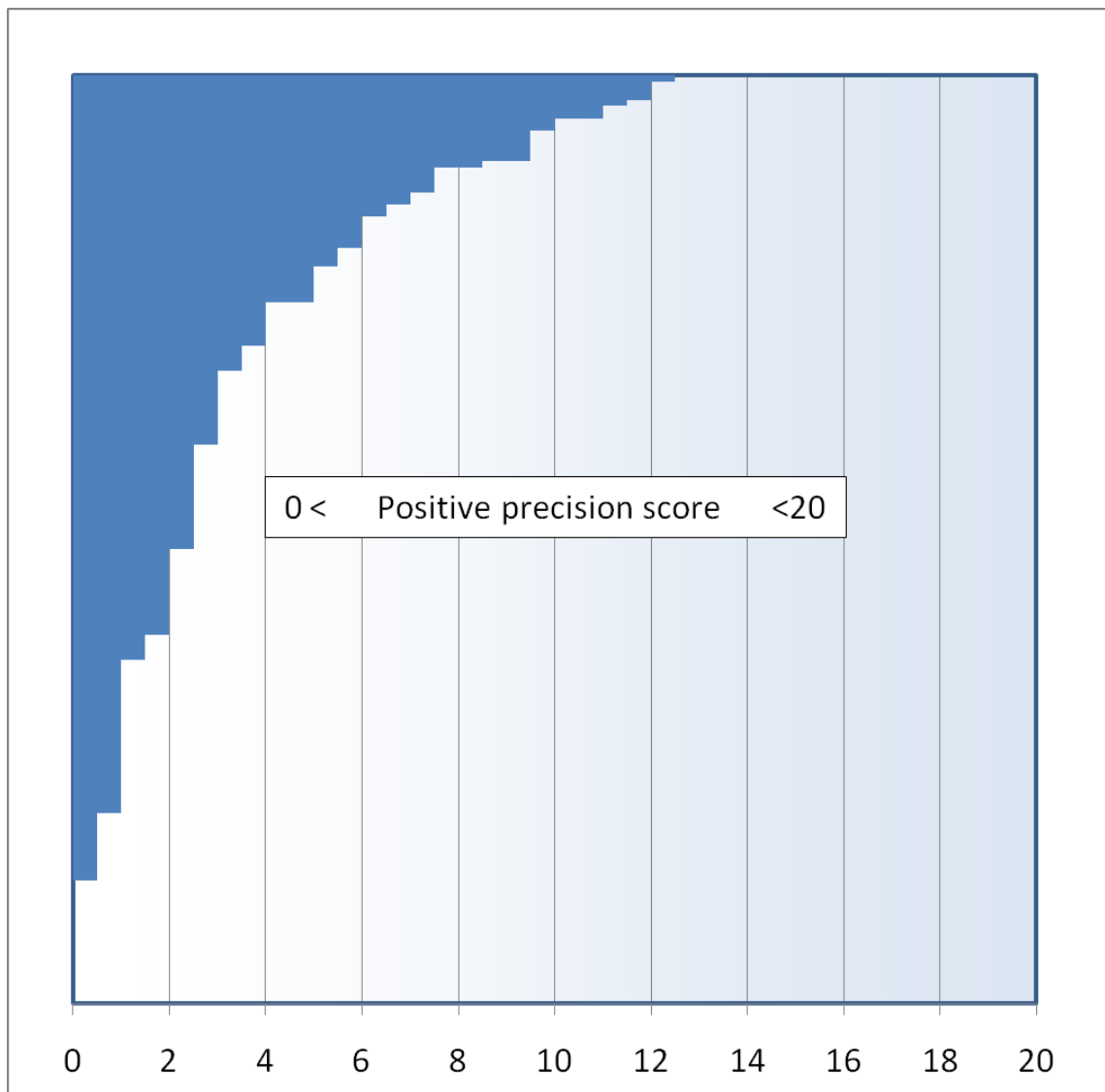


Figure 2. Positive precision scores for queries in WWW image searching.

Misinformation and the WWW

In the scatter chart in Figure 3, each symbol corresponds to one set of query results and shows the total positive precision score as well as the total negative misinformation score. This shows again that the precision is far from ideal. Furthermore the chart shows that misinformation cannot be neglected.

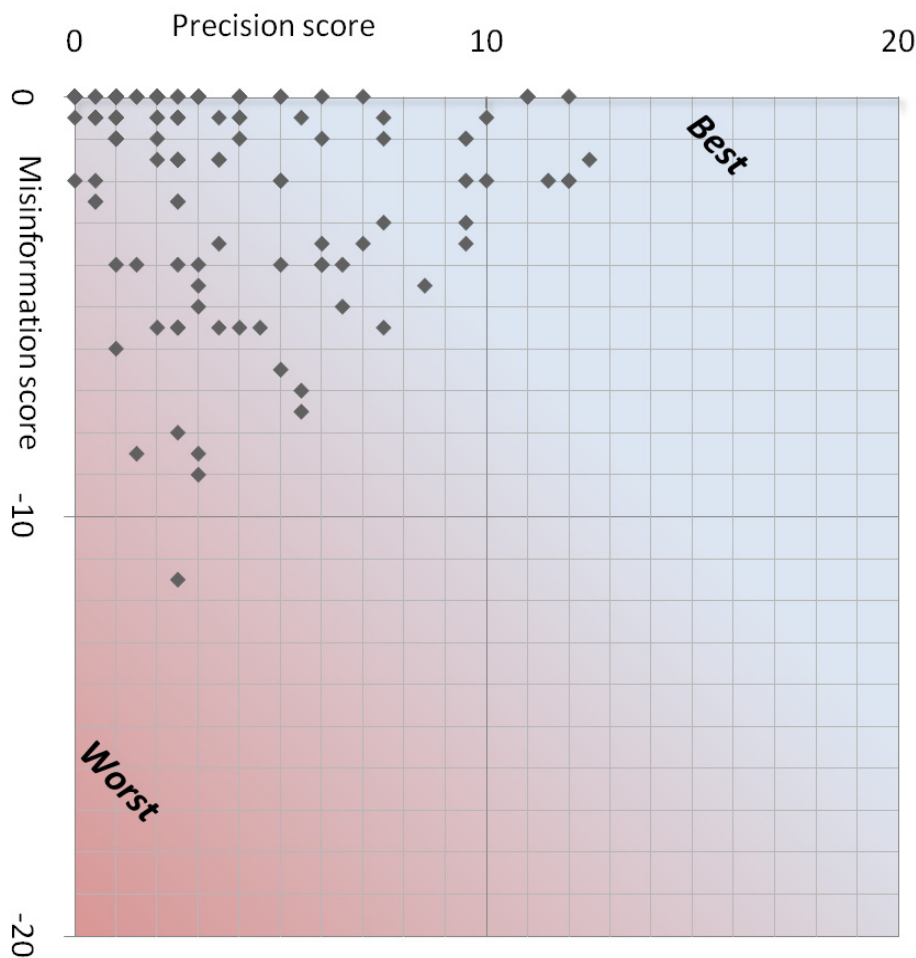


Figure 3. Total positive score for precision and total negative score for misinformation, for each image query result set of 20 images.

Here we can take into account that an individual image in a search result set of 20 images cannot be evaluated as completely relevant (with a positive score) and at the same time as completely misinforming (with a negative score). In a more mathematical, quantitative formulation, each retrieved image receives a score in such a way that (the positive relevance score) + (the absolute value of the negative misinformation score) is maximum = 1. So for a result set of 20 retrieved images (the sum of 20 positive relevance scores) + (the absolute value of the sum of 20 negative misinformation scores) is maximum = 20. Thus in the square chart with a square plot area, half of that area cannot be populated and is indeed not populated. Therefore the chart has been changed to a less common form with a plot area in the shape of a triangle, in Figure 4.

Figure 4 shows that the negative misinformation scores are significant in comparison with the positive precision scores, at least in the subject domain studied. In other words, many of the image results cause the user to lose time and may even lead to misinformation, particularly when a user is not an expert in the subject domain.

The horizontal line is drawn to divide the plot area in two parts: in the upper part each symbol can be interpreted as a net positive search experience, while in the lower part each symbol can be interpreted as a net negative search experience.

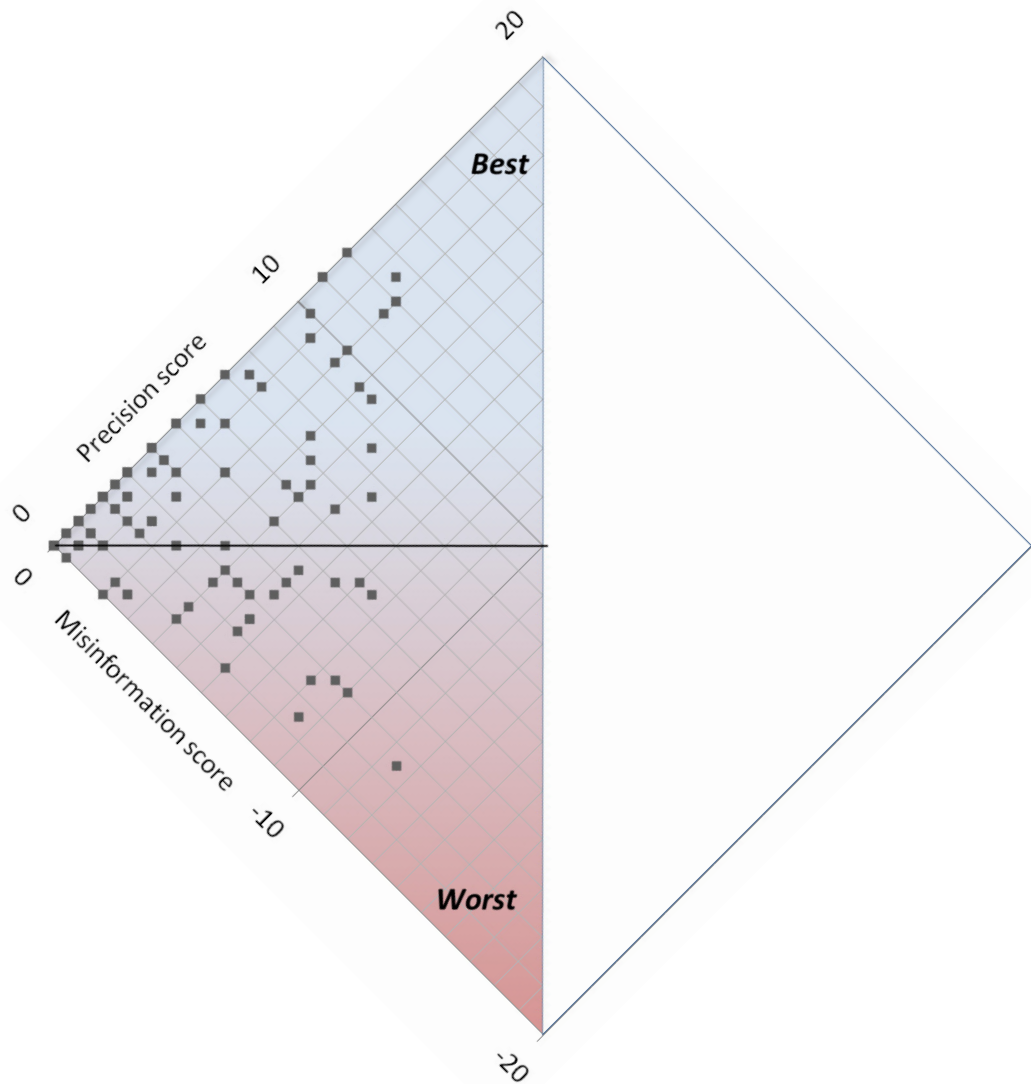


Figure 4. For each image query result set of 20 images, the chart shows the total positive score for precision and the total negative score for misinformation

Misinformation in More Detail

The positive as well as the negative score that has been assigned to each search result set have also been visualized on 2-dimensional scatter charts in which an additional variable was taken into account, namely the market value of the type of objects that was the focus of the search query. To realize this in practice, market values were assigned as explained in the section Methods above. When the highest observed market value is brought in the chart, we obtain Figure 5.

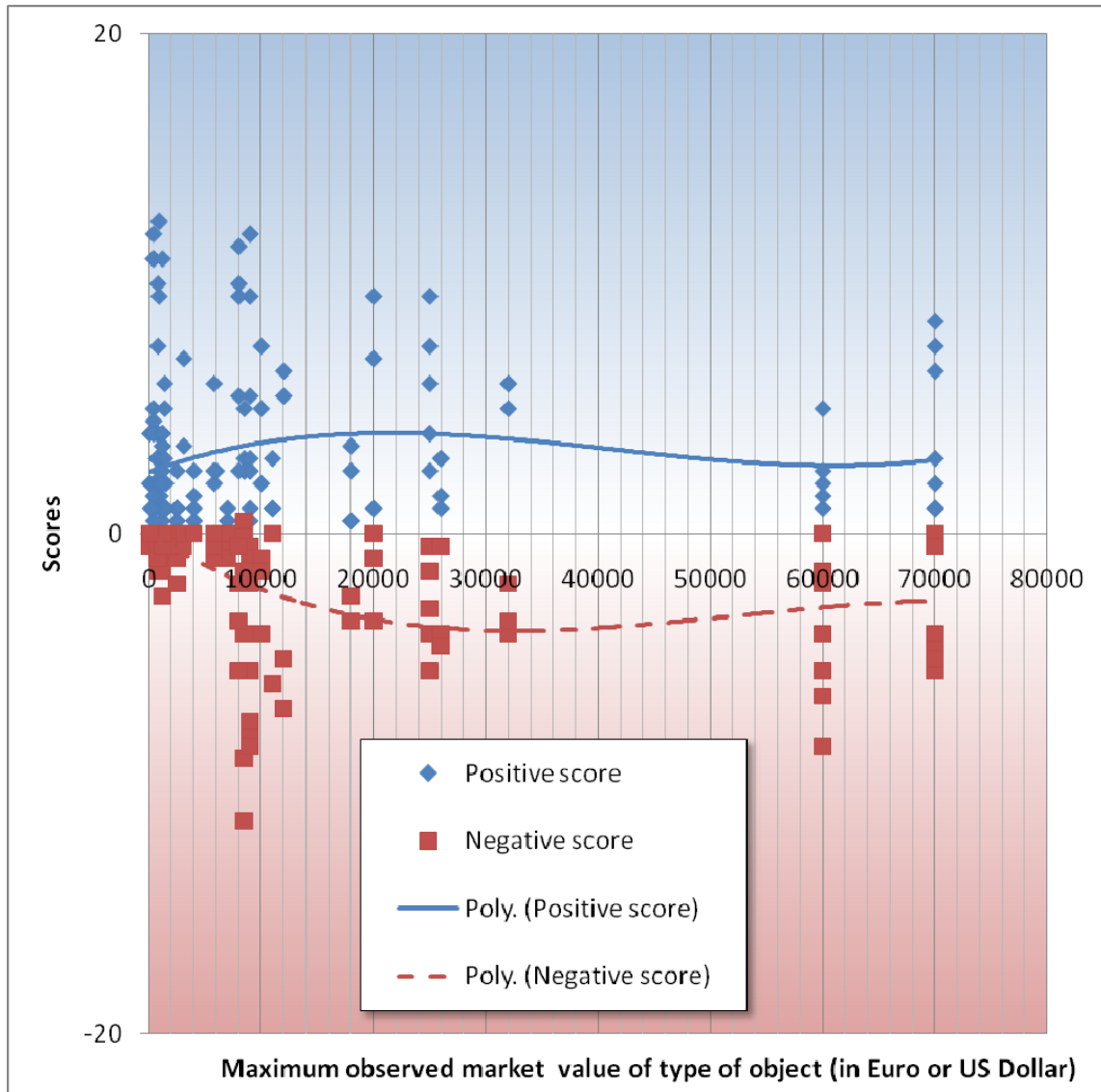


Figure 5. Total scores for each result set of a search on a type of African art object, in relation to the maximum market value of such an object

At least this chart gives the reader an idea about the market values of the objects discussed here. As expected, strong correlations between scores and market value are not observed, at least not at first sight. The two trendlines that are shown also have been calculated as polynomials from the data.

Most of the objects have a maximum observed market value that is low in comparison with the most valued objects, so that the right part of the chart is hardly populated and exploited. Taking this into account, we have improved the chart by using a logarithmic scale for the maximum observed market values to obtain Figure 6.

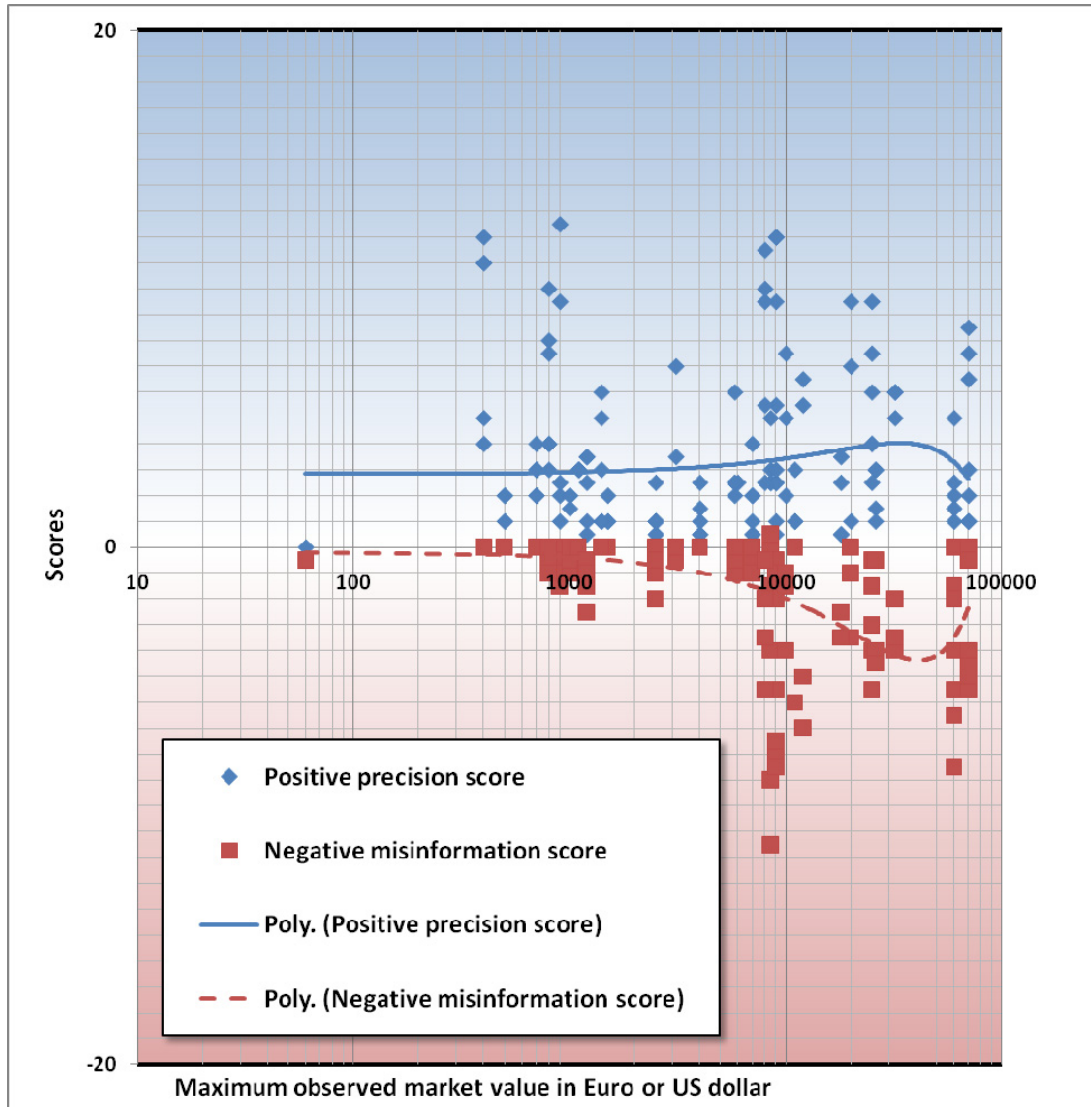


Figure 6. Total scores for each result set of a search on a type of African art object, in relation to the maximum market value of such an object.

By using a logarithmic scale for the maximum market value observed, the chart gives us a better view on the data collected for objects with a relatively low market value. This reveals that for objects with a maximum observed market value under 1000 Euro or US Dollar, most of the negative scores are much closer to zero than for objects with a higher market value. This is also reflected in the form of the trendline for the negative misinformation score: this starts from zero and goes down for market values higher than 1000 Euro or US Dollar.

Choosing the maximum observed market value of a type of objects to bring market value in the analysis is just one approach. A disadvantage is that the maximum is by definition an extreme and this can be exceptional, not reflecting an average situation. To eliminate at least partly the occurrence of extreme exceptions in this analysis, we have also taken into account the 2nd highest observed market value, as described in the section above on Methods. This results in an analogous

chart in Figure 7. As a consequence of choosing the other indicator of market value in the horizontal dimension, each symbol is shifted to a lower horizontal position.

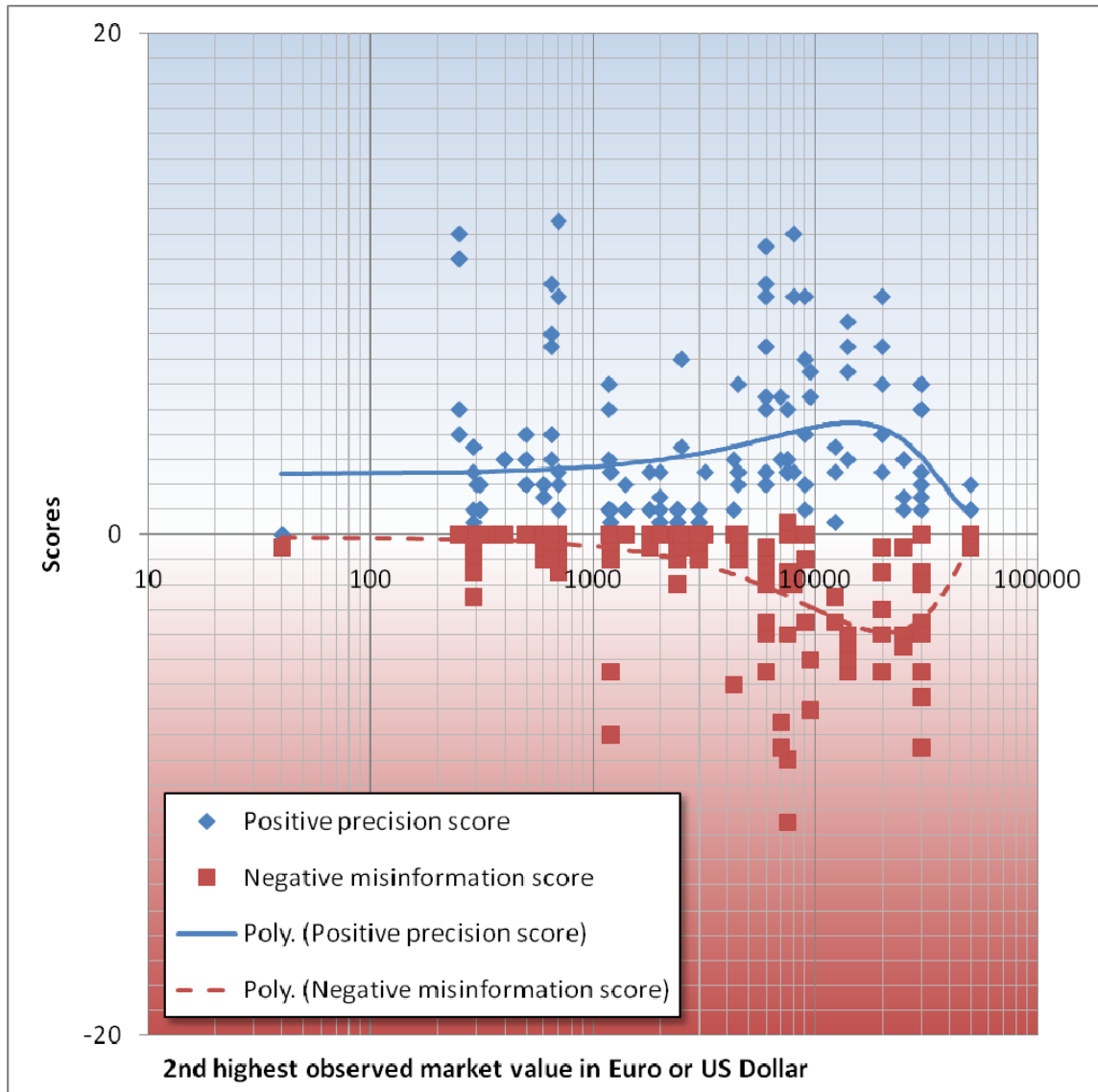


Figure 7. Total scores for each result set of a search on a type of African art object, in relation to the market value of such an object.

This chart leads essentially to the same conclusion: misinformation is less a problem for objects with a relatively low market value, more concretely for those types for which the maximum observed market value is less than 1000 Euro or US Dollar. This range of objects includes:

- metal objects that were used in a specific area as currency and/or anklet,
- organic calabashes that were used as container,
- textiles used as cloths,

- wooden headrests,
- wooden slingshots, etc.

Such functional objects receive attention in this field of art as their design and form is often considered as attractive and artistic. The following refer to some examples:

- A famous collection dedicated to top quality African functional objects has been exhibited and documented (Ginzberg, 2000). This went under the flag “African forms” and the word “art” was less prominent. Part of this collection has later been sold and distributed in an auction by a prestigious international auction house (Sotheby’s, 2007).
- Traditional African currencies have been collected by a bank and were exhibited in Madrid (Catálogo de la exposición Forma y valor. Monedas africanas tradicionales, 2008, and http://www.fundacionbancosantander.com/actividades_ficha.cfm?idArticulo=796).

As the popularity and market value is relatively low, copies or other derivations are not (yet?) spread over the world. This explains at least partly the findings in this analysis.

This outcome of the analysis was not expected in advance and is quite pleasing. Some more data collection seems desirable and may be a worthy time investment.

It is preferable to use in this analysis not the maximum but the 2nd highest observed market value, as this can at least partly exclude excessive exceptions. Therefore we go on with the data in the previous chart and display them in a different way in the next chart, Figure 8. There, each positive score is combined with the corresponding negative score to give an effective net score. These net scores summarize all observations for each search result, concerning precision and misinformation. In this way they give us a simple concentrated indication of what we can call effectiveness of one search action.

The chart in Figure 8 gives us a purified, minimalistic view, simpler than the previous charts and leads to the following findings:

- In general, the net score of all searches is much lower than the ideal value (here 20), due to a precision that is lower than the ideal and made even lower due to the occurrence of what we call misinformation on the WWW.
- For a significant number of searches, the net scores are even negative due to low precision and significant misinformation.
- More specifically, the net score is on average close to 3/20 for types of objects with a low market value and decreases further when the search subjects are types of objects with a higher market value.
- Even more detailed, the net scores are nearly all positive for result sets that correspond with searches for types of object with a market value lower than 1000 Euro or US Dollar. An explanation for this is that the WWW is nearly free of misinformation for these objects, as written in more detail above.

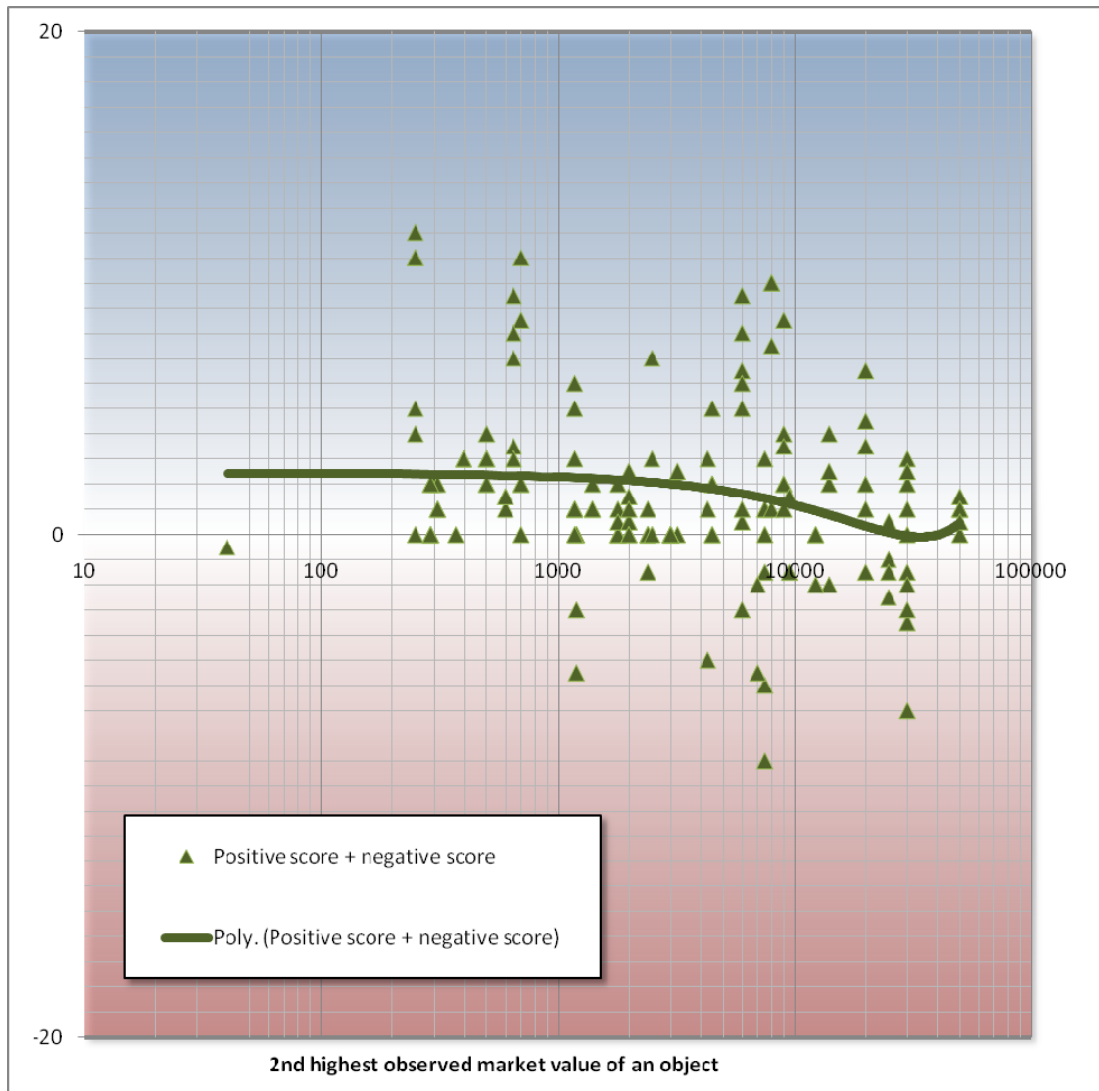


Figure 8. Total net score for each result set of a search on a type of African art object, in relation to the market value of such an object.

Discussion

The results shown are not a subjective selection from many tests, but an objective report of all tests, trying not to introduce any bias in the analysis.

“Operational definitions of quality criteria are needed.” This is one of the conclusions of Eysenbach et al. (2002) who reviewed “Empirical studies assessing the quality of health information for consumers on the World Wide Web”. Also Ostry et al. (2007) write that it may be important to think about developing practical and effective rating tools in order to guide consumers through the maze of often unreliable information on the WWW.

On the one hand I agree that quality criteria are needed or at least desirable, in view of the lack of a standard scoring method, so that the scoring procedure above had to be invented and refined during the work reported here.

On the other hand, scoring criteria and procedures should and will depend on the domain and aims of the investigation that involves scoring. Furthermore confusion is possible here, as the object of a scoring can be:

1. the contents of a document or
2. the search engine that allows discovery of documents or
3. the synergy of both as seen in the real-world digital media of these days.

This explains at least partly why standardization is absent in this kind of investigations, as far as I have seen.

Many retrieved images link to sellers of objects made for the trade that have almost no relation with more authentic objects. This is of course misleading. Concretely for this kind of subjects, users can get the impression that African art is not or never of a high level. People can come to the same conclusion when they see so called African art only in decoration shops, hotels, resorts, airports, shops in Africa and so on. So the misinformation observed in WWW searching is a reflection of the misinformation seen in real life. In reality we can see scarce high quality objects as well as many low quality fake, grotesque objects and all kinds of varieties between these extremes.

Noise in the form of low quality objects can perhaps be avoided by excluding results from the Internet domain named .com, as many low-end commercial dealers have their WWW site in that domain. However, Google Image Search or the other search engines do not seem to offer a simple way to achieve this, not even in the “advanced” search mode.

On the one hand, optimists can interpret the general results of this investigation as positive, since in many cases some relevant information is found easily and fast. On the other hand, critical pessimists can interpret the results as negative or at least as deserving some improvement in the future, because:

- most information that is publicly accessible in principle on the WWW is not discovered;
- most information that is publicly accessible in principle in printed published books is not found;
- precision is far below 100%, even though the quality of the queries was high;
- perhaps worse, many search results lead to commercial and other sites that deliver even misleading information.

Is the importance and occurrence of misinformation in the search results surprising?

On the one hand YES, because present day search engines base their relevance ranking of results not only on analysis of text in the documents on the WWW, but also on the number of links received from other WWW sites, and even on the quality of those links as evaluated by some algorithm (Manning et al., 2008). So this should yield a better ranking based on quality of content as assessed / judged by authors of other documents on the WWW, who decide to include links to external WWW sites in their own documents.

On the other hand NO, because highest quality objects are mainly shown by museums and other not-for-profit organizations. Due to their nature, such organizations spend probably less attention and efforts and funds to search engine optimization of their WWW site than more commercial for-profit organizations.

Conclusion and Recommendations

Conclusion

Information retrieval from the WWW through so-called image searching is attractive, simple and fast, but the outcome is still far away from the ideal that may be expected by some users:

- Users should be aware of the fact that the recall of relevant information sources is small in comparison with the documentary information that is in principle available.
- Users can be frustrated, lose time and can lose interest due to the low precision of result sets.
- Users can be misinformed, as many sources that are present on the WWW have not been reviewed and approved by experts, in contrast with most sources in printed documents or in scholarly digital publications.

Recommendations

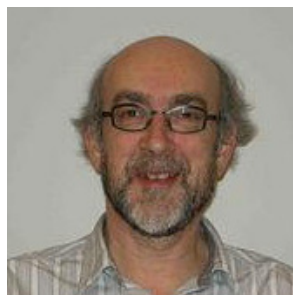
1. Authors, publishers and librarians can enrich the WWW with information that is hidden in valuable printed documents up to now and with additional valuable born digital information.
2. All types of librarians and educators should assist in upgrading the information literacy skills of all kinds of users. As an expected result, end-users can work more efficiently and more critically in the real, imperfect information world in general and in particular in the domain investigated here.
3. Users of information retrieval systems and information sources on the WWW, ranging from information experts to inexperienced end-users, should take into account the following:
 - Image search engines can be quite useful besides WWW text search engines and other retrieval systems.
 - Unfortunately, in real, practical cases like the ones investigated here, WWW image searching yields a low precision.
 - The difficulty or danger of misleading links and information sources on the WWW should be emphasized, and is probably higher than in the case of more classical, printed published information sources. So a critical attitude is appropriate.

References

- Bacquart, J-B. (2002). *The tribal arts of Africa: Surveying Africa's artistic geography*. Thames and Hudson Ltd., 240 pp. [first edition published in 1998]
- Bassani, E. (2004). *Afrique: Aux origines de l'art moderne* [exposition]. Turin: GAM Galleria d'Arte Moderna e Contemporanea, Skira Editore.
- Bassani, E. (2005). *Arts of Africa: 7000 years of African art*. Monte Carlo: Grimaldi Forum, Skira Editore, 412 pp.
- Bouttiaux, A-M. (2009). *Persona. Masks of Africa: Identities hidden and revealed*. Milano: 5 Continents, 304 pp.
- Carini, V. (2004). *A hidden heritage: Sculture africane in collezioni private italiane*. Milano: Galleria Somare, 344 pp.
- Catálogo de la exposición "Forma y valor. Monedas africanas tradicionales."* (2008). Madrid: Fundación Banco Santander y Fundación Alberto Jiménez- Arellano Alonso, 10.04.2008 > 31.05.2008, ISBN: 978-84-89913-89-9, 98 pp.
- De Mul, J. (2009). The work of art in the age of digital recombination. In M. van den Boomen, S. Lammes, A-S. Lehman, J. Raessens, & M. T. Schafer (Eds.), *Digital material – Tracing new media in everyday life and technology* (pp. 95-106). Amsterdam: Amsterdam University Press, ISBN 978 90 8964 068 0, e-ISBN 978 90 4850 666 8. Available online from:

- [http://www2.eur.nl/fw/hyper/Download/De%20Mul%20\(2008\)%20The%20Work%20of%20Art%20in%20the%20Age%20of%20Digital%20Recombination.pdf](http://www2.eur.nl/fw/hyper/Download/De%20Mul%20(2008)%20The%20Work%20of%20Art%20in%20the%20Age%20of%20Digital%20Recombination.pdf)
- Eisenhofer, S., & Guggei, K. (2002). *Afrikanische Kunst: Fakten, Preise, Trends (Weltkunst Antiquitäten-Fuhrer)*. Munich, Berlin: Deutscher Kunstverlag, 163 pp.
- Eysenbach, G., Powell, J., Kuss, O. & Sa, E-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *JAMA*, 287, 2691-2700.
- Ginzberg, M. (2000). *African forms: The traditional design and function of objects*. [exhibition catalogue with foreword by Jack Lenor Larsen] Milan : Skira Editore, 298 pp.
- Grootaers, J-L., & Eisenburger, I. (Eds.). (2002). *Forms of wonderment*. [2 volumes] Berg en Dal: Afrika Museum.
- Hahner, I., Vajda, L., & Kecskesi, M. (2007). *African masks: The Barbier-Mueller collection*. Prestel, ISBN 3791338072, 290 pp. [Published also earlier in a larger format version.]
- Herreman, F., Burssens, H., Chadeisson, M., Cole, H. M., Dewey, W. J., Foss, P., et al. (2003). *Material differences: Art and identity in African Art* [Exhibition catalogue]. New York: Museum for African Art, and Gent: Snoeck-Ducaju & Zoon, 180 pp.
- JISC. (2008). *Review of image search engines*. Retrieved from <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/review-of-image-search-engines/>
- Kamer, H. (1974). De l'authenticité des sculptures Africaines = The authenticity of African sculptures. [paper in 2 languages]. *Arts de l'Afrique noire*, 12, 17-40.
- Maertens, H., Neerman, M., & Burssens, H. (2009). *African Faces: Een hommage aan het Afrikaanse masker*. Uitgeverij Lannoo [2008. First edition; 2009 Second, larger edition with some content removed and some content added].
- Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press, 482 pp.
- Martínez-Jacquet, E. [in collaboration with de Bérénice Geoffroy-Schneiter]. (2009). *Regards de marchands. La passion des arts premiers*. [Exposition]. Monnaie de Paris – 09 septembre / 18 octobre 2009, Editeur: Primedia (Belgique), 290 pp.
- Miller, J. (2006). *Tribal art* [DK Collector's Guides]. DK Publishing, 240 pp.
- Ostry, A., Young, M. L., & Hughes, M. (2007). The quality of nutritional information available on popular websites: A content analysis. *Health Education Research*, 23(4), 648-655.
- Petridis, C. (2003). *South of the Sahara: Selected works of African art*. Cleveland Museum of Art, 128 pp.
- Pierrat, E. (2008). *Comprendre l'Art africain*. Editions du Chêne, 310 pp.
- Robbins, W. M., & Nooter, N. I. (2004). *African art in American collections* (2nd ed.). Atglen, PA: Schiffer Publishing.
- Sotheby's. (2007). *The Marc and Denyse Ginzberg Collection: African forms*. [catalog], Paris, 10 Septembre 2007, 166 p. Available online with auction results from http://www.sothebys.com/app/live/lot/LotResultsDetailList.jsp?event_id=28504&sale_number=PF7027
- Stepan, P. (2001). *World art: Africa*. Prestel Publishing, 128 pp.
- Stepan, P. & Hahner, I. (2005). *Spirits speak: A celebration of African masks*. Munich: Prestel Publishing, 190 pp.
- Visona, M., Poynor, R., Cole, H. M., & Blier, S. P. (2007). *A history of art in Africa*. London: Prentice Hall, 560 pp.
- Weinhold, U. (2001). *Het eeuwige gezicht. / The eternal face. Afrikaanse maskers en de westerse samenleving. African masks and Western Society*. Berg en Dal: Afrika Museum, 234 pp.

Biography



Since 1983, **Paul Nieuwenhuysen** is a full-time member of the academic staff at the Vrije Universiteit Brussel, nowadays as professor. These days his functions include: member of the management board of the University Library, science and technology librarian, as well as teaching courses on online information retrieval and presentation. At the University of Antwerp inter-university postgraduate program in Information and library science, he was guest professor responsible for courses on information technology and on the information market until 2009.

At the University of Antwerp he received the degrees of Licentiaat in Physics in 1974, Doctor in Science in 1979, the Belgian post-doctoral degree (named Geaggregeerde voor het Hoger Onderwijs) in 1983, and the inter-university postgraduate degree in Documentation and library science in 1986.

He has organized international training courses on management of information in science and technology and has been project leader of a co-operation with the National Agricultural Library of Tanzania 1998-2008.

He is author or co-author of more than 30 refereed publications in international scientific/technical journals, he is a member of several societies and of the editorial board of journals in the area of information science, and he has been a consultant for various international agencies.

Appendix

The following are examples of concrete, specific search queries used in this investigation. These should not be confused with the real information needs / concepts / topics that form the basis, the starting point for these queries and that cannot be summarized in just a few words; those topics are not described here, simply due to a lack of space here in this paper.

yoruba slingshot / lobi maternite / lobi maternity / lobi mother child / kirdi cache sex / kirdi cache sexe / kirdi apron / kirdi schamschutz / kirdi skirt / kirdi dance / poto knife / poto sword / poto sword african / poto sword ceremonial / poto knife ceremonial / topoke / liganda / liganda topoke / lokele / lokele liganda / mbuti / mbuti tapa / mbuti barkcloth / mbuti pongo / pygmy barkcloth / mossi / mossi mask / mossi bousla mask / kwele money / kwele currency / kwele money currency / kambata / kambatta / kambatta neckrest / kambatta neck rest / kambatta headrest / mfunte money / mfunte money currency / mfunte currency / tusian mask / toussian mask / toussian plank mask / loniaken / tousian mask / trumbash ivory / trombash ivory / trombash knife / trombash knife ivory / mangbetu knife ivory / mangbetu knife / pende / gitenga / pende gitenga / senufo pulley calao / senufo pulley / senoufo pulley / senoufo pulley calao / senoufo / senufo hornbill / senufo hornbill pulley / dagari / dagari bangala / dan gameboard / dan game board / dan mancala / dan mankala / dan spiel / dan spiel afrika / namji / namji doll / namji doll 2 / namchi / namchi doll / dowayo / dowayo doll / boa mask / boa mask congo / boa masque / boa masque congo / kuba bwoom / bwoom / bwoom mask / bakuba bwoom / akuaba / ashanti akuaba / mossi doll / biiga / mossi biiga / bambara janus / bamana janus / kplekple / goli kplekple / mwana / mwana mask / mwana pwo / ntomo / ntomo mask / ntomo bamana / ntomo bambara / ntomo mali / ntomo masque / salampasu mask / mukinka / salampasu / salampasu mask copper / hembra / hembra singiti / singiti / lwalwa / lwalwa mask / lwalwa mfondo / lwalwa mvondo / lwalwa mkaki / mvondo / tiv / tiv nigeria / tiv Ihambe / tiv nigeria Ihambe / Ihambe / kanaga / kanaga mask / kanaga mali / kanaga masque / kanaga dogon / satimbe / satimbe mask /

satimbe masque / satimbe dogon / afo / afo nigeria / afo money / afo currency / afo
currency nigeria / afo galma / afo galma currency / mama mask / mama nigeria / mama
nigeria mask / mama mask bushcow / mumuye / mumuye mask / mumuye vabo / mu-
muye masque / vabo mask / vabo masque / mwana hiti / mwana hiti zaramo / mwana
hiti kwere / mwana hiti zaramo kwere / mbole / mbole currency / mbole currencies /
mbole monnaie / mbole money / mongo / mongo currency / mongo money / mbole en-
kelband / mbole legband / mbole fussreif / mbole copper / mbole bronze